

## **Pendeteksian Keberfungsian Butir Diferensial Soal USBN Kimia SMA Negeri di Kabupaten Bone dengan Metode Khi Kuadrat Lord**

Nursaida Hardiyanti  
Universitas Negeri Makassar, Makassar  
[nursaidaht@gmail.com](mailto:nursaidaht@gmail.com)

**Abstract:** *This research aims at discovering bias items of USBN test for Chemistry Subject if detected through Lord Chi-Square based on gender bias which held at public high schools in Bone Regency, Academic Year of 2016-2017. This uses exploratory-descriptive design with qualitative approach of Item Response Theory. The samples of this research include all twelfth grade students at public high schools in Bone Regency as much as 676 test participants. The data source are taken from students' answer sheets at public high schools in Bone Regency. The technique of data analysis in this research is quantitative analysis using R program with three-parameter model of item response theory and detecting method of Lord Chi-Square. To sum up, the findings show that by using Lord Chi-Square, there are 19 items detected as bias based on gender which significantly contain DIF (Differential Item Functioning).*

**Keywords:** *Dif, gender, IRT, Lord's Chi Square method*

**Abstrak:** Penelitian ini bertujuan untuk mengetahui Butir-butir tes USBN kimia yang memuat bias butir bila dideteksi dengan metode Khi kuadrat Lord berdasarkan bias gender SMA Negeri di Kabupaten Bone tahun pelajaran 2016/2017. Penelitian ini merupakan penelitian deskriptif - eksploratif dengan pendekatan kuantitatif dengan Pendekatan Teori Respon Butir. Subjek penelitian ini adalah seluruh peserta didik kelas XII IPA SMA Negeri di Kabupaten Bone sebanyak 676 peserta tes. Sumber data dalam penelitian ini adalah lembar jawaban peserta didik kelas XII IPA SMA Negeri di Kabupaten Bone. Teknik analisis data menggunakan analisis kuantitatif menggunakan program R dengan pendekatan teori respon butir model 3 parameter dengan menggunakan metode pendeteksian Khi Kuadrat Lord. Hasil penelitian ini menunjukkan bahwa dengan menggunakan metode Khi Kuadrat Lord ada 19 butir terdeteksi bias berdasarkan gender yang secara signifikan memuat DIF.

**Kata Kunci:** *DIF, gender, IRT, metode Khi Kuadrat Lord*

## 1. Pendahuluan

Evaluasi adalah suatu proses yang sistematis untuk menentukan atau membuat keputusan sampai sejauh mana tujuan-tujuan pengajaran telah dicapai oleh siswa. Salah satu bentuk pengukuran siswa terhadap hasil belajar adalah Ujian Sekolah Berstandar Nasional (USBN). USBN adalah kegiatan pengukuran capaian kompetensi siswa yang dilakukan sekolah untuk mata pelajaran tertentu dengan mengacu pada standar kompetensi lulusan untuk memperoleh pengakuan atas prestasi belajar. Dengan demikian instrument tes yang disusun untuk USBN seharusnya memiliki validitas dan reliabilitas yang baik. Tes merupakan alat yang digunakan sebagai sarana untuk menentukan nilai. Perlu memperhatikan validitas soal, reliabilitas soal, derajat kesukaran item, daya pembeda item, fungsi pengecoh, serta kesesuaian soal dengan standar kompetensi dan kompetensi dasarnya agar kita dapat menentukan kualitas soal.

Menurut Djemari, Kumaidi & Badrun (2011), untuk memberikan penilaian yang adil, instrumen penilaian harus bebas dari unsur bias butir tes yang disebabkan adanya *differential item functioning* (DIF)". Perangkat soal USBN akan menjadi sesat jika tidak objektif, adil, dan memihak kelompok tertentu sehingga dua siswa dari kelompok yang berbeda dengan kemampuan yang sama mempunyai peluang menjawab benar yang berbeda. Perangkat soal tersebut menjadi tidak baik kualitasnya karena mengandung sifat bias pada butir-butirnya. Oleh karena itu, pendeteksian terhadap ada tidaknya sifat bias pada

tes dan butir-butir tes USBN sangat perlu untuk dilakukan agar ketidakadilan atau kerugian pada kelompok tertentu dapat dihindarkan dan kemampuan siswa terukur secara objektif.

Salah satu syarat soal memiliki fungsi informasi tinggi adalah yang tidak mengandung unsur bias di dalamnya sehingga dapat menyebabkan kelompok siswa yang memiliki kemampuan yang sama, memiliki peluang menjawab soal yang bebrbeda. Atau dalam artian bahwa soal atau tes itu memiliki karakteristik butir yang baik sehingga mampu membedakan kemampuan karakteristik responden sehingga tidak merugikan salah satu pihak tertentu.

Menurut Lord (1980) mengemukakan bahwa "jika pada suatu tes memuat butir-butir yang memihak pada kelompok tertentu, maka tes tersebut dikatakan memuat bias atau mengandung keberfungsian tes diferensial (*Differential Test Functioning*)". "Selain keberfungsian butir diferensial, untuk mengindikasikan tes adil bagi semua golongan, dapat digunakan keberfungsian tes diferensial (*DTF*)" Stark & Chernyshenko (dalam Retnowati, 2006) .

Selain itu, pendeteksian DIF dapat memberikan informasi mengenai butir soal yang akan digunakan dalam suatu tes. Namun pada kenyataannya masih banyak guru-guru disekolah yang kurang menyadari pentingnya pendeteksian bias dilakukan untuk mendapatkan soal yang benar-benar berkualitas. Jika hal ini dibiarkan tanpa ditanggulangi maka sistem evaluasi yang digunakan oleh para guru sebagai penentu keputusan akan memiliki permasalahan. Jadi perlu ada terobosan

yang bisa menanggulangi permasalahan yang telah dialami oleh para pendidik selama ini. Sehingga siswa mendapat keadilan dari perolehan prestasi yang sesuai dengan kemampuannya. Selain dari kesalahan yang terjadi pada penyusunan instrument, “perbedaan jenis kelamin juga dapat menimbulkan bias butir. Bias tes yang paling sering muncul adalah pada perbedaan kelompok pada jenis kelamin, perbedaan ras dan etnis (Hambelton dan Rogers :1995).

Selanjutnya apabila ditinjau dari segi materi pada mata pelajaran di sekolah, konsep materi pada mata pelajaran kimia adalah ilmu yang mencari jawaban atas pertanyaan apa, mengapa, dan bagaimana gejala-gejala alam yang berkaitan dengan komposisi, struktur dan sifat, perubahan, dinamika, dan energetika zat. Termasuk pada mata pelajaran kimia di SMA yang merupakan mata pelajaran yang mempelajari segala sesuatu tentang zat yang meliputi komposisi, struktur dan sifat, perubahan, dinamika dan energetika zat yang melibatkan keterampilan dan penalaran. Topik-topik kimia harus dipelajari dengan urutan tertentu. Oleh sebab itu, pembelajaran kimia dan penilaian hasil belajar kimia harus memperhatikan karakteristik ilmu kimia sebagai proses dan produk. Untuk itu tes pada mata pelajaran kimia sangat memungkinkan mengandung bias khususnya jika ditinjau berdasarkan gender dan wilayah.

Pada penelitian ini bias butir soal dideteksi dengan pendekatan teori respon butir. Kim & Cohen (1995) menyatakan bahwa ada tiga metode untuk mendeteksi bias dengan pendekatan teori respons butir, yaitu : (1) khi kuadrat menurut Lord, (2)

metode luasan menurut Raju, dan (3) uji perbandingan kemungkinan (*Likelihood Ratio Test*), dan pada penelitian ini menggunakan metode khi kuadrat menurut Lord. Penggunaan metode ini bertujuan untuk menggambarkan berapa banyak butir soal USBN yang dapat dideteksi bias berdasarkan gender.

Metode Lord's  $\chi^2$  yaitu membandingkan peluang menjawab soal kedua kelompok pada suatu item dengan memeriksa perbedaan parameter item tergantung pada model IRT yang digunakan (Lord, 1980). Selanjutnya Lord (1980) dalam Price, E.A. (2014) menyatakan bahwa metode Lord, membandingkan parameter kelompok fokus dan parameter kelompok referensi, melihat apakah perbedaan kedua kelompok signifikan secara statistik. Selanjutnya menurut Price, E.A. (2014) bahwa pengujian ini untuk kesetaraan ICC antara kelompok referensi dan kelompok fokus. Manfaat penggunaan metode ini dapat mendeteksi DIF konsisten (Uniform) dan tidak konsisten (Non-Uniform).

Camili & shepard (dalam Retnawati, 2014) menyatakan bahwa “untuk menguji hipotesisnya, digunakan uji statistik  $X^2$  yang berderajat kebebasan 3”. Selanjutnya menurut Anne.D.S.(2014) Item diidentifikasi sebagai DIF jika memiliki Chi-square b-DIF yang signifikan statistik pada  $p < 0,05$ .

Menurut Lopez, G.E. (2012), menyatakan bahwa metode DIF Lord, dapat digunakan untuk menguji perbedaan dalam satu atau lebih parameter secara bersamaan di seluruh kelompok referensi dan kelompok fokus. Kelebihan metode ini mudah diadaptasi ke model parametrik

apapun, nilai kritisnya mudah diperoleh untuk df dan tingkat signifikansi yang berbeda, dan indeks sensitif terhadap keduanya DIF konsisten dan tidak konsisten.

Analisis uji kecocokan model dilakukan untuk melihat kesesuaian perangkat soal dengan tiga model parameter logistik teori respon butir. Uji kecocokan model yang akan digunakan adalah ANOVA dengan melihat nilai AIC (Akaike Information Criteria). AIC digunakan untuk mengatasi permasalahan pemilihan model. Model yang terbaik adalah model dengan skor AIC terendah (Snipes & Taylor, 2014). Adapun model logistik yang dipilih untuk penelitian ini yaitu model logistik 3P.

Hasil penelitian Mansyur dan Muliana (2016), tentang Pendeteksian DIF dan DTF ujian akhir sekolah mata pelajaran matematika, menunjukkan bahwa dari 40 butir soal terdeteksi 10 butir soal terindikasi bias butir (DIF) berdasarkan gender dan 13 butir terindikasi bias berdasarkan lokasi (wilayah) sedangkan yang terdeteksi memuat bias tes (DTF) terdiri dari 6 butir bias tes tidak konsisten dan sebanyak 6 butir bias tes konsisten berdasarkan gender yang cenderung menguntungkan perempuan.

Hasil penelitian Louis (2016), tentang mengukur Item Recovery (Pemulihan Butir) dan Keberfungsian Diferensial Butir untuk Model Logistik Tiga Parameter, menunjukkan bahwa pada pengujian skala besar, penggunaan pendekatan ragam kelompok (multigroup) dibatasi dalam pengukuran keberfungsian diferensial butir (DIF) antar variabel berganda sebagai DIF yg diuji tiap variabelnya secara terpisah. Sebaliknya, prosedur teori respon butir dengan kovariat

(IRT-C) dapat digunakan untuk mengukur DIF antara variabel berganda (kovariat) secara bersamaan. Untuk mengukur kegunaan prosedur IRT-C, dapat menerapkan penelitian simulasi. Dengan menggunakan data SAT (Scholastic Assessment Test)/ujian standar untuk masuk perguruan tinggi sebagai parameter realistis, DIF beraturan pada tiga kovariat disimulasikan pada: jenis kelamin (dikotomi), ras/etnis (kategori), dan pendapatan (secara berkala). Simulasi telah dilaksanakan pada beberapa kondisi berikut: dua jenis lama tes (14 dan 21 butir soal), empat contoh sampel (5,000, 10,000, 20,000, 40,000), dan dua ukuran efek DIF (sedang, besar). Telah ditemukan bahwa prosedur IRT-C dapat memulihkan nilai rata-rata yg tersembunyi serta pengukuran model logistik tiga parameter secara tepat dengan ukuran sampel besar yaitu 20,000.

Hasil penelitian Efendi (2011), tentang Pendeteksian *Crossing Differential Item Functioning* (CDIF) Berdasarkan Teori Respon Butir, menunjukkan bahwa ketiga metode dapat digunakan untuk pendeteksian CDIF, tetapi metode luasan menurut Raju sangat sensitif dalam mendeteksi CDIF di bandingkan dengan kedua metode yang lain. Disamping itu tidak ada DIF yang disebabkan perbedaan gender pada tingkat kemampuan tinggi.

Hasil penelitian Retnawati (2003), tentang Pendeteksian keberfungsian butir diferensial pada perangkat tes seleksi masuk di tiga SMP di Kota Yogyakarta. Hasil penelitian menunjukkan hasil tes tersebut menguntungkan kelompok perempuan pada seluruh parameter peserta tes

## 2. Metode Penelitian

Jenis penelitian ini merupakan jenis penelitian eksploratif deskriptif dengan pendekatan kuantitatif yakni akan mengungkap bias butir pada perangkat soal USBN pada mata pelajaran kimia dengan menggunakan metode pendeteksi bias berdasarkan teori respons butir yaitu metode khi kuadrat Lord. Data dalam penelitian ini adalah respon peserta USBN kimia yang diperoleh dari SMA Negeri di Kabupaten Bone tahun pelajaran 2016/2017. Sumber data yang digunakan yaitu berupa lembar jawaban siswa yang telah didokumentasi.

Subjek dalam penelitian ini adalah seluruh peserta USBN di SMA Negeri di Kabupaten Bone tahun pelajaran 2016/2017. Pemilihan subjek dilakukan dengan menggunakan teknik purposif sampling. Peneliti menentukan ukuran sampel yaitu sebanyak 676 respon peserta didik yang tersebar pada beberapa sekolah tersebut.

Teknik pengumpulan data yang dilakukan dalam penelitian ini adalah dengan menggunakan teknik dokumentasi, yaitu dengan mengutip respon peserta USBN Kimia di SMA Negeri di Kabupaten Bone tahun pelajaran 2016/2017.

Pendeteksian bias butir pada penelitian ini menggunakan pendekatan teori respon butir dengan menggunakan metode khi kuadrat Lord. Metode khi kuadrat Lord ini akan melihat persamaan dengan vektor dari perbedaan pengestimasi parameter item, untuk mendeteksi bias butir menggunakan metode khi kuadrat Lord, digunakan persamaan adalah sebagai berikut.

$$X^2 = v_i \sum^{-1} v_i$$

Dengan  $v_i$  merupakan vektor dari perbedaan pada pengestimasi parameter item ke- $i$  antara grup fokal dan grup referensi  $\sum^{-1}$  merupakan matriks varians kovarians untuk perbedaan pada pengestimasi parameter item (Camilli & Shepard, 1994).

## 3. Hasil Penelitian dan Pembahasan

**Deskripsi Hasil Penelitian Berdasarkan Differential Item Functioning (DIF) dengan menggunakan Metode Khi Kuadrat Lord berdasarkan gender**

Berdasarkan hasil analisis data dengan metode difLord diperoleh informasi bahwa dari 35 butir soal diperoleh 19 butir soal yang terindikasi memuat bias gender. Pada penelitian ini, suatu butir terdeteksi bias butir dengan metode difLord apabila nilai  $\chi^2_{hitung}$  lebih besar dari  $\chi^2_{tabel}$  dengan signifikansi 0,05. Perhitungan nilai  $\chi^2$  ini dengan bantuan program R. Dalam hal ini program R menghendaki taraf signifikansi terdeteksi sebagai dif adalah kurang dari 0,05.

Adapun hasil analisis bias gender dengan metode difLord diperoleh informasi bahwa terdapat 19 butir yang memuat bias gender yaitu butir 2, 3, 8, 9, 12, 14, 15, 16, 18, 20, 21, 24, 25, 26, 27, 30, 31, 32, dan 35. Hasil ini dibuktikan dengan nilai signifikansi yang kurang dari 0,05. Sesuai dengan pendapat Anne D.S. (2014) yang menyatakan bahwa item diidentifikasi sebagai DIF jika memiliki chi-squared b-DIF yang signifikan statistik  $p < 0,05$ .

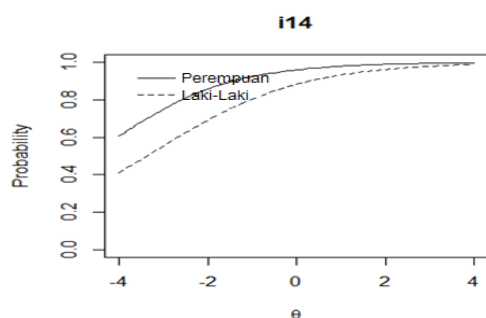
**Tabel 1.** Hasil Analisis Bias Gender Soal USBN SMAN Negeri di Kabupaten Bone pada Mata Pelajaran Kimia Tahun Pelajaran 2016/2017

Butir	Stat.	P-value	Keterangan	Butir	Stat.	P-value	Keterangan
i1	0.7912	0.6733		i19	0.6784	0.7123	
i2	45.6596	0.0000	***	i20	8.3244	0.0156	*
i3	44.6176	0.0000	***	i21	8.3642	0.0153	*
i4	0.4387	0.8031		i22	2.0089	0.3662	
i5	0.643	0.7251		i23	3.4101	0.1818	
i6	3.5551	0.169		i24	166.1755	0.0000	***
i7	0.024	0.988		i25	7.2612	0.0265	*
i8	41.037	0.0000	***	i26	58.5012	0.0000	***
i9	57.8023	0.0000	***	i27	90.9769	0.0000	***
i10	4.861	0.088	.	i28	1.203	0.548	
i11	4.9448	0.0844	.	i29	0.7724	0.6797	
i12	71.8036	0.0000	***	i30	55.0257	0.0000	***
i13	5.2189	0.0736	.	i31	61.3085	0.0000	***
i14	10.8183	0.0045	**	i32	26.6676	0.0000	***
i15	8.5507	0.0139	*	i33	2.9273	0.2314	
i16	11.4499	0.0033	**	i34	2.5312	0.2821	
i17	4.4006	0.1108		i35	20.6377	0.0000	***
i18	15.7448	0.0004	***				

Selanjutnya, dilihat dari grafik ICC pada butir yang terdeteksi bias, dapat ditentukan kelompok yang diuntungkan dan kelompok yang dirugikan oleh butir yang deskriminan

Adapun dari 19 butir yang terdeteksi bias tersebut, terdapat dua jenis bias yang terjadi yaitu bias secara tidak berpotongan/konsisten (*uniform*) dan bias secara berpotongan/konsisten (*non uniform*). Berikut beberapa grafik perbedaan antara butir-butir yang terdeteksi bias *uniform* dan bias *nonuniform*

a) Bias *Uniform* (grafik bias butir 14 dan butir 31

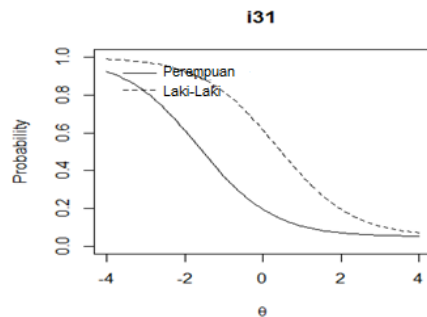


**Gambar 1** Kurva DIF pada Soal USBN Tahun Pelajaran 2016/2017 SMAN di kabupaten Bone.

Gambar 1 menunjukkan bahwa grafik tersebut tidak saling berpotongan artinya termasuk dalam kategori DIF konsisten (*Uniform*). DIF *Uniform* terjadi jika pada grafik tersebut tidak berpotongan pada salah satu titik dan terjadi pada setiap skala level kemampuan (*ability*).

Gambar 1 di atas menunjukkan bahwa pada kemampuan  $\theta = 1$ , didapat probabilitas menjawab butir soal dengan benar kelompok laki-laki (*focal*) dan kelompok perempuan (*reference*) berbeda. Artinya bahwa kemampuan peserta laki-laki dan perempuan sama dengan 1, tetapi probabilitas menjawab butir soal dengan benar yang tidak sama pada butir 14, sehingga butir tersebut terdeteksi sebagai DIF.

Untuk grafik butir soal 31 juga termasuk dalam bias *Uniform*.

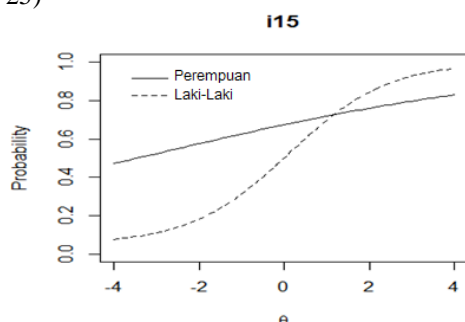


Gambar 2 Kurva DIF secara keseluruhan pada Soal USBN Tahun Pelajaran 2016/2017 SMAN di kabupaten Bone.

Gambar 2 menunjukkan bahwa grafik tersebut tidak saling berpotongan artinya termasuk dalam kategori DIF konsisten (*Uniform*). DIF *Uniform* terjadi jika pada grafik tersebut tidak berpotongan pada salah satu titik dan terjadi pada setiap skala level kemampuan (*ability*).

Gambar 2 di atas menunjukkan bahwa pada kemampuan  $\theta = 0$ , didapat probabilitas menjawab butir soal dengan benar kelompok laki-laki (*focal*) dan kelompok perempuan (*reference*) berbeda. Artinya bahwa kemampuan peserta laki-laki dan perempuan sama dengan 0, tetapi probabilitas menjawab butir soal dengan benar yang tidak sama pada butir 31, sehingga butir tersebut terdeteksi sebagai DIF.

b) Bias *Uniform* (grafik bias butir 15 dan butir 25)

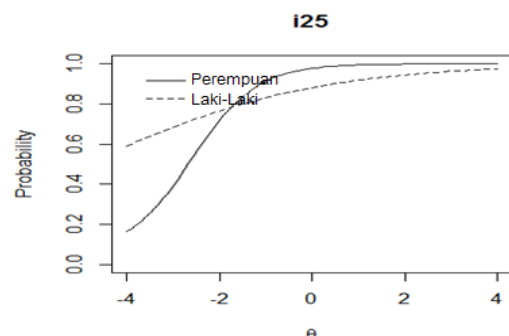


Gambar 3 Kurva DIF secara keseluruhan pada Soal USBN Tahun Pelajaran 2016/2017 SMAN di kabupaten Bone

Gambar 3 menunjukkan bahwa grafik tersebut saling berpotongan, di level kemampuan sekitar 1 artinya termasuk dalam kategori DIF tidak konsisten (*Non Uniform*). DIF *Non Uniform* terjadi jika pada grafik tersebut berpotongan pada salah satu titik dan tidak terjadi pada setiap skala level kemampuan (*ability*).

Gambar 3 di atas menunjukkan bahwa pada kemampuan  $\theta = 2$ , didapat probabilitas menjawab butir soal dengan benar kelompok laki-laki (*focal*) dan kelompok perempuan (*reference*) berbeda. Artinya bahwa kemampuan peserta laki-laki dan perempuan sama dengan 2, tetapi probabilitas menjawab butir soal dengan benar yang tidak sama pada butir 15, sehingga butir tersebut terdeteksi sebagai DIF.

Untuk grafik butir soal 25 juga termasuk dalam bias *Non uniform*



Gambar 4 Kurva DIF secara keseluruhan pada Soal USBN Tahun Pelajaran 2016/2017 SMAN di kabupaten Bone.

Gambar 4 menunjukkan bahwa grafik tersebut saling berpotongan, di level kemampuan sekitar -1,8 artinya termasuk dalam kategori DIF tidak konsisten (*Non Uniform*). DIF *Non Uniform* terjadi jika pada grafik tersebut berpotongan pada salah satu titik dan tidak terjadi pada setiap skala level kemampuan (*ability*).

Gambar 4 di atas menunjukkan bahwa pada kemampuan  $\theta = 0$ , didapat probabilitas menjawab butir soal dengan benar kelompok laki-laki (*focal*) dan kelompok perempuan (*reference*) berbeda. Artinya bahwa kemampuan peserta laki-laki dan perempuan sama dengan 0, tetapi probabilitas menjawab butir soal dengan benar yang tidak sama pada butir 25, sehingga butir tersebut terdeteksi sebagai DIF.

Berdasarkan hasil analisis data menggunakan metode khi kuadrat Lord dengan menghitung nilai  $\chi^2$  dan membandingkan dengan tabel nilai kritik  $\chi^2$  dengan derajat kebebasan dan menggunakan taraf signifikansi yang dikehendaki. Dalam hal ini program R menghendaki taraf signifikansi memuat DIF adalah kurang dari 0,05. Adapun butir yang nilai  $\chi^2$  kurang dari 0,05 untuk bias gender berjumlah 19 butir. Butir yang terdeteksi DIF adalah butir butir 2, 3, 8, 9, 12, 14, 15, 16, 18, 20, 21, 24, 25, 26, 27, 30, 31, 32, dan 35. Hal ini sejalan dengan pendapat Anne D.S. (2014: 64) yang menyatakan bahwa item diidentifikasi sebagai DIF jika mereka memiliki chi-squared b-DIF yang signifikan statistik  $p < 0,05$ .

Selanjutnya pendapat Kim dan Cohen (1995) bahwa bila  $\chi^2$  lebih besar dari taraf signifikansi 0,05 maka butir itu mengandung DIF. Pendapat lain dari Raju (1993), menyatakan bahwa prosedur dalam mendeteksi DIF dalam konteks teori respon butir *Lord's chi-square test*, mengidentifikasi item yang memuat DIF signifikan secara statistik.

Gender yang dimaksud dalam penelitian ini adalah kelompok laki-laki yang disimbolkan sebagai kelompok *focal* dan kelompok perempuan yang disimbolkan kelompok *reference*

Selanjutnya, dilihat dari ICC pada butir yang terdeteksi bias, dapat ditentukan kelompok yang diuntungkan dan kelompok yang dirugikan oleh butir yang deskriminan.

Berdasarkan hasil pendeteksian bias, pada bias gender dari 19 butir soal yang terdeteksi bias terdapat 8 butir soal yang menguntungkan kelompok laki-laki (*focal*), serta terdapat 11 butir soal yang lebih menguntungkan kelompok perempuan (*reference*). Hal ini berarti bahwa soal ini lebih menguntungkan peserta tes kelompok perempuan dan merugikan kelompok laki-laki.

Sebagaimana yang dijelaskan oleh Djemari, Kumaidi & Badrun (2011) mengemukakan “untuk memberikan penilaian yang adil, instrumen penilaian harus bebas dari unsur bias butir tes yang disebabkan adanya differential item functioning (DIF)”. Karena bias ini akan menyebabkan adanya unsur diskriminasi pada salah satu kelompok peserta tes.

Hal ini didukung oleh pendapat Camili & Sherpard (1994 : 62) mengemukakan bahwa “sebuah butir dikatakan tidak bias jika probabilitas keberhasilan pada butir soal tersebut adalah sama untuk mengambil tes yang memiliki kemampuan sama dari populasi yang sama tanpa memperhatikan keanggotaan kelompok mereka”.

Kedua pendapat di atas sejalan dengan (Crocker & Algina, 1986 : 268) yang mengemukakan “Bias butir adalah perbedaan probabilitas menjawab benar butir soal dari dua kelompok yang berbeda setelah mengontrol tingkat kemampuan”. Artinya bias dapat diketahui penyebab dan sumbernya setelah kemampuan telah diestimasi, dengan terlebih dahulu



mengontrol kelompok yang diinginkan oleh kemampuan sesuai dengan kecurigaan terlebih dahulu.

Berdasarkan beberapa pendapat para ahli di atas sehingga dapat disimpulkan bahwa bias butir adalah kesalahan persepsi salah satu kelompok peserta tes yang memiliki kemampuan sama dalam memaknai stem soal yang disebabkan oleh beberapa hal dengan terlebih dahulu mengestimasi kemampuan berdasarkan kecurigaan sebelumnya.

#### 4. Kesimpulan

Berdasarkan hasil analisis data dan pembahasan penelitian yang telah dikemukakan dan dihubungkan dengan rumusan masalah, maka peneliti berhasil mendeteksi 19 butir soal yang terindikasi bias gender yaitu butir Soal USBN mata pelajaran Kimia di SMA Negeri di Kabupaten Bone yaitu 2, 3, 8, 9, 12, 14, 15, 16, 18, 20, 21, 24, 25, 26, 27, 30, 31, 32, and 35 yang cenderung menguntungkan perempuan.

#### 5. Daftar Pustaka

- Anne, D.S. 2014. Accuracy Of Differential Item Functioning Detection Method In structurally Missing Data Due To Booklet Design. *Dissertation*. The University British Of Columbia
- Camilli, G., & Shepard, L. A. 1994. *Methods fo Identifying Biased Test Items*. Thousand Oaks, CA: sage Publication.
- Crocker, L & James A. 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart, and Winston.
- Djemari, Kumaidi, & Badrun, 2011. Pengembangan Instrumen Pengukur Hasil Belajar Nirbias dan Terskala Baku. *Jurnal Penelitian dan Evaluasi Pendidikan*, jilid 15, No. 2
- Effendi.2011. Pendeteksian Crossing Differential Item Functioning (Cdif): Berdasarkan Teori Responsi Butir.. *Jurnal Evaluasi Pendidikan* .Vol. 2 No. 2
- Hambleton, R. K., Swaminathan H., & Rogers, H. J., 1992. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publication Inc.
- Kim, Seock-Ho& Cohen, Allan. 1995. Detection of differential item functioning in multiple groups. *Journals of Educational Measurement, Fall95, Vol. 32 Issue 3*
- Lopez, G.E. 2012. Detection and Classification of DIF Types Using Parametric and Nonparametric Methods: A comparison of the IRT-Likelihood Ratio Test, Crossing-SIBTEST, and Logistic Regression Procedures. *Dissertation*. Department of Psychology College of Arts and Sciences University of South Florida
- Lord, F. M., 1980. *Applications Of Item Response Theory To Practical Testing Problems*. New Jersey, CA: Lawrence Erlbaum Associates, Publiher Hillsdale.

Mansyur dan Muliana.2016.Detecting Differential Item Functioning and Differential Test Functioning On Math School Final Exam. *International Journal On Advanced Science Engineering Information Technology*. Vol 6 No.4

Price, Emily A.2014. Item Discrimination, Model Data FIT, Type I Error Rates in DIF Detection Using Lord's  $\chi^2$  the Likelihood Ratio Test, and the Mantel Haenszel Procedure. *Dissertation*. The Patton College Of Education of Ohio University.

Retnawati, H. 2003. Keberfungsian Butir Diferensial pada Perangkat Tes Seleksi Masuk SLTP Mata Pelajaran Matematika. *Jurnal Penelitian dan dan Evaluasi Pendidikan*. Nomor 6.

Snipes & Taylor. 2014. Model selection and Akaike Information Criteria: An example from wine ratings and prices. *Artikel*. University of Houston

Tay, Louis dkk.2016. Item Response Theory With Covariates (IRT-C): Assessing Item Recovery and Differential Item Functioning for the Three-Parameter Logistic Model. *Educational And Psychological Measurement*. SAGE. Vol 76 (1).